

Using the PhonePass SET-10 Test to Measure Oral English Ability

Catlin Hanna

Growing interest in communicative language teaching, emphasizing meaningful interaction rather than knowledge of linguistic rules, has been accompanied by interest in descriptions of functional language ability and methods of measuring oral proficiency. Turner (1998) found that the publication of the ACTFL (American Council on the Teaching of Foreign Languages) Guidelines and the creation of the ACTFL Oral Proficiency Interview (1986) marked the beginning of more than a decade of investigation in oral proficiency interview tests, including how they compared to other test formats and exploration of techniques for testing oral skills in specific educational settings.

In Japan, the current competitive and rapidly changing work environment has made English communication skills a valuable asset. Students who graduate from college or university with only minimal, survival-level oral skills will be at a disadvantage in many fields. In fact, graduates may feel that they have “wasted their time” in English courses if they see no noticeable improvement in their ability to use the language.

Fortunately, English courses with the goal of increasing communicative competency are now part of most college and university language programs. At the same time, gauging their effectiveness is not a straightforward matter.

In an effort to evaluate the general effectiveness of the English language instruction in the Department of Intercultural Studies, a two-year junior college program, the ITP TOEFL exam was introduced. Beginning in 1998, it has been given to the first year students one month after their entrance into the department and again at the end of their second and final year in the program. During the four years that the exam has been administered, the students' mean scores have risen an average of 18 points from the first to the second administration of the test. Some students' scores rose as much as 40 points and some dropped slightly over the two years.

While these results are largely positive, it is not clear how the TOEFL measures — reading, grammar, and listening — reflect assessment of students' speaking competence. Our program features small conversation classes taught by native speakers, yet it is difficult for instructors to get an objective picture of students' ability to communicate in English from observing the students in class and the ITP TOEFL alone.

It was the desire to have a more complete picture of the oral ability of students that led the department to investigate tests of spoken English and to administer the PhonePass SET-10 test to the first-year students in June 2002. Our interest in the attendant issues and our experience with the PhonePass test will be addressed below.

Speaking Tests

Researchers investigating spoken English testing have noted the many inherent difficulties in testing oral language ability. Testing speaking as a discrete skill is almost impossible and of questionable validity. Success in speaking depends in part on the listener as well, which makes tester reliability and objectivity an issue.

Genuine talk is a spontaneous matter, to which a listener spontaneously responds. It is not clear how a scripted test situation reflects the ability of a student in spontaneous talk. On a practical level, it can be difficult, time consuming and costly to test large numbers of students under identical conditions (Kitao,1996). The need, however, to determine the candidates' ability to communicate in English has led to the development of a variety of exams both overseas and in Japan.

The most well known of the overseas tests include direct speaking tests such as the Cambridge Assessment of Spoken English (CASE), the Interagency Language Roundtable (ILR) Oral Proficiency Interview (OPI), the American Council on the Teaching of Foreign Languages (ACTFL) test, and the Educational Testing Service's (ETS) Test of English for International Communication (TOEIC). The TOEIC interview test can only be taken by those who are among the top two of five ranks in the preliminary written multiple-choice test, but otherwise there is close similarity among these interview tests. These interview tests have high face validity because they require the examinees to use spoken language. Their drawbacks include the need for trained native speakers as raters, and the fact that they are not designed to adequately discriminate among lower level students (Nakamura,1994).

There are also semi-direct speaking tests such as the Test of Spoken English (TSE) and SPEAK test, developed by Educational Testing Service (ETS), and the Association of Recognized English Language Schools Examinations Trust's Oral Examinations (ARELS Oral Examinations). In these exams, the candidates' answers are recorded on tape and evaluated by trained raters. The questions are either printed or recorded. The criticism leveled at this type of exam is that interaction with a tape is not authentic communication. (Nakamura,1994) However, the test can be conducted under standardized conditions at one time to a large number of people and the answers from a variety of tasks can help the rater construct a detailed picture of the examinee's ability without any visual distracters.

In Japan, the Society for Testing English Proficiency (STEP), founded in 1963, developed six grades of tests for assessing Japanese candidates' English proficiency. Many students however, do not take the "speaking" test because they are not able to pass the preliminary multiple-choice written exam. With the six grades, candidates are able to set individual goals regarding the level they would like to reach. (<http://www.eiken.or.jp/english/>) However, training for examiners is not well established and it cannot be said to be a standardized test from a statistical point of view (Nakamura,1994). In fact, the Japanese Ministry of Education, Culture, Sports, Science and Technology which is now "developing a strategic plan to cultivate "Japanese with English abilities"" (<http://www.mext.go.jp/english/>) ended its long-standing connection with STEP in 2001, a move likely to result in a decrease in its popularity and perceived validity in Japan.

Research to determine the validity of semi-direct exams in comparison to direct, interview-type tests has found that although semi-direct exams relying on a tape recording are less desirable, there is a high degree of correlation between scores of both types of tests, and that semi-direct tests can be considered a "reasonable alternative" to direct testing when the latter is impractical (Clark and Swinton,1980 ; Stansfield,1990).

PhonePass SET 10 Exam

A recently developed alternative is the PhonePass test, made available in Japan two years ago. Developed by the U.S.-based Ordinate Corporation, the PhonePass SET-10 test is a ten-minute telephone test that uses computer software to measure spoken English. Before the test, the examinees are given test sheets with pin numbers, an explanation of the different sections of the test, and the sentences they will be instructed to read. The examinees are free to check words they are unfamiliar with and practice reading the sentences prior

to the taking the test. To take the test, the examinee makes a telephone call and gives an identifying pin number. The examinee is then led through the four sections of the exam by a recorded voice.

In Part A of the test, the candidates are instructed to read eight of the twelve sentences on the paper they have been given.

In Part B, the examinee listens to and repeats a series of 16 sentences.

In Part C, which also consists of 16 items, the examinee hears a word and says a word with the opposite meaning.

In Part D, the examinee gives short answers to 16 questions.

In the final section of the test, Part E, the examinee hears an open ended question, has eight seconds to consider a response and thirty seconds to answer. Part E is not scored by PhonePass, but the responses are recorded and available for the faculty administering the test to listen to when they access the scores on the Ordinate website.

The SET-10 provides a score report consisting of an overall score and four sub-scores. The overall score represents the ability to understand English and "speak it intelligibly at a native conversational pace on everyday topics". Sub-scores measure performance in sentence mastery, vocabulary, fluency, and pronunciation. All scores are reported on a scale of 2 to 8 ([http : //www.ordinate.com](http://www.ordinate.com)).

According to the Ordinate Corporation website, the test uses a speech recognition system which "includes an HMM-based speech recognizer using acoustic models, pronunciation dictionaries, and expected-response networks that were custom-built from data collected during administrations of PhonePass tests to over 400 native speakers and over 3500 non-native speakers of English" (Validation Summary for PhonePass Set-10, [http : //www.ordinate.com](http://www.ordinate.com), October,2002).

Results

Fifty-three students took the PhonePass SET-10 exam over two weeks in June and July of 2002. Ten students who initially received no score because they did not speak loudly enough or because their answers were indecipherable were retested. Two of the ten received no score the second time and were not tested again. Scores ranged from 5.1 to 2.3, with the mean score being 3.4.

On first viewing, the PhonePass scores appeared to vary a bit from the ITP TOEFL scores. But a scatterplot of the scores of both tests (see Fig.1) illustrates a strong positive correlation ($r=.81$) between results of the two tests. High correlations between the ITP TOEFL and the PhonePass SET-10 exam were noted in the "Validation Summary for PhonePass SET-10" (www.ordinate.com, October,2002) and suggest that both are testing the same general language ability.

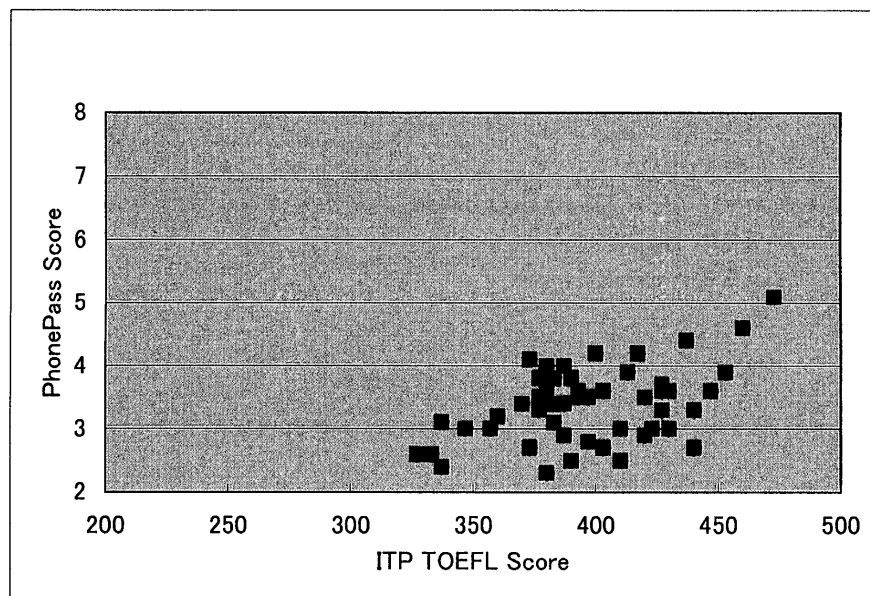


Figure 1. Scatterplot of students' ITP TOEFL Scores and PhonePass Scores reflecting a correlation of $r=.81$.

While students with high scores on one test tended to have high scores on the other (and vice versa), as the scatterplot graph shows, there were students with the same ITP TOEFL score but quite different PhonePass scores. This supports the notion that while “tapping” the same general language ability, the two tests are also directly evaluating different skills as they were designed to do. Thus, for example, students with the same level of proficiency in written English may have different levels of proficiency in oral English.

Questionnaire

After all students completed the PhonePass test, we gave them a short questionnaire consisting of four questions in Japanese asking for their reactions to the test.

The first question asked which section of the test was most difficult and why. Twenty-five students responded that Part B (Repeat) was the most difficult because the sentences were too long, and spoken too quickly to understand and repeat. Twenty-four students responded that Part E (Open Questions) was the most difficult. The reasons given were that they were unable to understand clearly what was being asked, and so were not able to answer and also did not have enough time to consider how to answer. Five students responded that Part D (Questions) was the most difficult section because they were not able to understand the questions.

The second question asked whether taking the test and receiving their score had made the student feel more motivated to study English, less motivated, or the same as before taking the test. Thirty-four students said they felt more motivated, 14 felt the same, and 2 felt less motivated.

In the third question, students were asked whether or not they would like to take the test again next year to see how their spoken English had improved. Forty-one students said that they would like to take the test again. Ten said that they were not interested in re-taking the test.

Question four asked the students what kind of English study they thought would help them improve their score on the test. Thirty-four students mentioned that listening to more English would be beneficial. Students mentioned such listening activities as tapes, music CDs, television dramas and news programs. Other

answers included increasing their English vocabulary (5 students), talking with native speakers such as the instructor (4 students), and working on pronunciation (2 students).

One student did not complete the questionnaire and some students did not answer all four questions. Some students may have felt that saying that the test made them more motivated to study and that they wanted to take the test again were the expected and “correct” answers. But clearly the “Repeat” and “Open Question” sections of the exam were the most challenging, with difficulty in listening preventing the students from answering well. This is consistent with what the English instructors noticed when proctoring the tests.

Discussion

The strengths of the PhonePass test and the reasons for it being selected by the department include its reliability, validity and convenience. According to research posted on the Ordinate website, PhonePass scoring is as accurate and reliable as a highly trained human grader. As well, there is a strong correlation between PhonePass scores and scores on three other widely used tests, the TOEFL (as indicated above), TOEIC and ILR Speaking Test (Validation Summary for PhonePass SET-10, [http : //www.ordinate.com](http://www.ordinate.com)).

Students receive a detailed explanation of the components of their score in English and Japanese and may feel that a test scored by a computer is more objective than an exam given by their own instructors. Also increasing its validity in the eyes of the students is the fact that the PhonePass exam, unlike an oral exam created and used only within the college, is likely to become popular among businesses and other organizations for hiring and promoting staff. Taking it twice in their college career, students will have a clear idea of their level and will be more prepared if they are later asked to take it “in the real world”.

Scores are available almost immediately after the test is taken and can be accessed by faculty with a password from any computer through the Ordinate website. Students can take the test from any phone at any time. In the case of the Department of International Cultural Studies, students took the test in the office of an English instructor. The ten-minute duration of the test makes it convenient for students, especially if they are taking the ITP TOEFL around the same time and might be overwhelmed by another long exam.

Speaking about the difficulty of creating an effective oral English exam, one of the PhonePass test’s developers, Prof. Jared Bernstein, says, “Previously it has been difficult to provide a standard measure of proficiency in spoken English that works with precision over a wide range of performance levels” (Harada and Bernstein, p.6,2002). He asserts that the SET-10 test provides an exam that is comparable over time and place, measures independent performance ability in listening, speaking and discussion, and provides highly consistent and reliable scores on a continuous interval scale covering a range of levels from middle school students to competent adult speakers (p.6).

Despite this claim, PhonePass, like a number of other English tests, may not be as effective at discriminating among fairly low-level examinees. Judging from student responses to the questionnaire and from overhearing the telephone tests, Part E, the “open questions” section, was extremely difficult. Three examples of Part E questions are :

“Do you prefer to read true stories about real people who have actually lived, or do you prefer stories about made-up people and events? Explain your choice.”

“Today, many people are living longer, and many old people are being cared for by their children or grandchildren. How does this situation affect the family?”

"If you were asked to represent your country at an international conference, what topic would you like to discuss with other people from other nations. Please explain why."

To answer after only eight seconds would be challenging even for native speakers and the necessity of having such difficult questions is unclear. A simple open-ended question could elicit a sophisticated response from a high-level candidate while giving lower-level examinees a chance to at least say something minimal. Most of the examinees did not answer at all because they felt that they did not understand the question sufficiently. This section is not scored; nevertheless, the students had no opportunity to demonstrate their ability to speak at any length in an unstructured way.

When asked about this, Stephen Black, the sales manager of PhonePass in Japan, said that even people whose scores are relatively high will still have a high percentage of questions that they cannot answer. In his experience "low-level students seem to be able to answer (correctly or incorrectly) about 30% of the questions, mid-level about 60% and higher around 90%" (personal communication, July 4, 2002)

Harada and Bernstein make the claim that PhonePass tests "discussion" skills in addition to speaking and listening (Harada and Bernstein, p.6, 2002). But it is unclear how English discussion skills are being tested in the PhonePass test since there is no actual back and forth talking taking place. The criticism that is made of other semi-direct tests can also be made of this test. Can speaking to a recorded voice on a phone that only captures the examinees' replies be considered a "discussion"? Even face-to-face oral interview exams have faced questions about whether they constitute "conversation" (Turner, 1998). The ability to negotiate meaning with another person in a dialogue or in a group discussion is a skill that most students find more difficult to master than speaking on their own, but it is a skill that is essential to be considered a competent speaker of a language.

Another possible weakness of the test is its cost of ¥4000 per test, making it almost twice as expensive as the ITP TOEFL. And although the test only takes ten minutes per student, the number of students who can be tested at one time is limited by the number of available phones that can be proctored. While it is possible for students to make the call on their own from any phone, the reliability of the scores would be lower since there is no way to verify that the students themselves actually made the call. In the case of the Department of Intercultural Studies, the four English instructors administering the test spent between two and five hours each answering student questions prior to the test and proctoring the phone calls of the 53 examinees.

Another difficulty encountered was students receiving no score because their voices were not loud enough or they were silent for too long. Both our students in this category appeared to the proctor to be speaking at normal volume. It may be that the voice recognition software has difficulty with some voices, but generally the problem was resolved with instructions to the students to speak clearly, hold the receiver correctly, and say "I don't know" when unable to answer a question.

Conclusion

Clearly more research will need to be done to determine both the effectiveness of the English language instruction in the Department of International Cultural Studies and the PhonePass test as a measure of students' oral English skills. In early 2003, the second year students will take the test for the first and only time. Their scores will provide some hint about the degree of improvement we might expect to see when the original test group takes the test a second time in 2004.

With this first, limited look, it appears that the PhonePass test would be effective for placing students in the type of English courses that emphasize oral skills. The test will also give some information about the degree

of improvement of students' spoken English over their two years in the program. However, the PhonePass test does not seem to provide the kind of complete picture of students' English communication skills that a more comprehensive test including a range of activities such as an interview, group discussion, and presentation would.

What can be said more certainly is that the current interest in investigating oral English assessment will have beneficial outcomes for both students and instructors. These may include including greater involvement of teachers in assessment, the integration of learning and assessment, and the development of clear, understandable rating criteria. Hopefully, the development of new, more convenient and reliable types of oral English assessment will have a powerful washback effect on English language education, making communicative language teaching the norm rather than the exception.

ACKNOWLEDGEMENTS

I would like to thank Prof. Christine Winskowski for her invaluable assistance with the scatterplot showing the correlation between student test scores.

References

- Clark, J. L. & Swinton, S. S.(1980) *The Test of Spoken English as a Measure of Communicative Ability in English-Medium Instructional Settings*. TOEFL Research Reports,7.
- Kitao,S.K., & Kitao, K.(1996) *Testing Speaking*, (ERIC Document Reproduction Service No. ED 398261)
- Harada, Y., & Bernstein, J.(2002,September) *Automatic Measurement of Spoken English Skills*. Paper presented at the JACET 41st Annual Convention, Tokyo, Japan.
- Nakamura, Y.(1994)An Examination of English Speaking Tests and Research on English Speaking Ability. Choufu Gakuen Women's Junior College Journal no.26.
- Stansfield,C.W.(1990)Some Foreign Language Test Development Priorities for the Last Decade of the Twentieth Century. Foreign Language Annals, v 23 395-401.
- Turner, J.(1998)At and Beyond Oral Proficiency Testing. Annual Review of Applied Linguistics, 180, 192-207.

Using the PhonePass SET-10 Test to Measure Oral English Ability

Catlin Hanna

Communicative language teaching, with its focus on functional language ability, is becoming more common in Japanese college and university English programs. It is often difficult, however, to gauge the effectiveness of the instruction in raising the level of oral English proficiency of students. Tests purporting to measure speaking skills do not always provide reliable and valid scores and can be prohibitively time consuming.

In an effort to gain a more complete picture of the English skills of its students, the Department of International Cultural Studies administered the PhonePass SET-10 exam (a speaking/listening exam administered by phone) in addition to the ITP TOEFL to the first year students in 2002. A strong correlation was found between the scores of the two tests, suggesting that they are reliable indicators of student ability. At the same time, individual differences in PhonePass scores among students with similar ITP TOEFL scores suggest that the PhonePass SET-10 may indeed “tap” students’ oral skills. Further testing will give a more complete picture of students’ progress in acquiring oral skills in comparison with their mastery of reading and grammar.